

8th International Conference on Information Quality (IQ-2002)

Barclays Bank Case Study: Using Artificial Intelligence to Benchmark Organizational Data Flow Quality

Adrian McKeon
Infoshare Limited
amckeon@infoshare-is.com

1

8th International Conference on Information Quality (IQ-2003)

Executive Summary/Abstract

- Most IT systems cannot measure the accuracy of outputs
 - ◆ Does the system work and where is the evidence?
 - ◆ Are business decisions based on garbage?
- Small data quality flaws at the start of a project magnify into inexplicable defects in end user outputs. Objectives cannot be translated into measurable performance indicators and nobody knows why.
- Barclays¹ used artificial intelligence to audit trail the history of each data record at sub field, field and record level from the source system to the warehouse. Audit trails made the workings of the IT system transparent and end users were able to validate output, identify errors and track back to fix them.

1 McKeon, A. J. "Basel II Compliance: Can You Prove It?" A Barclays Bank plc Case Study, November 2002, www.infoshare-is.com

2

8th International Conference on Information Quality (IQ-2003)

Which Comes First: Data or Technology?

- Why measuring data quality is important
- What happens when you can't measure it
- Barclays Bank Case Study
 - ◆ What did Barclays Achieve?
 - ◆ How did Barclays Achieve it?
- Conclusions

3

8th International Conference on Information Quality (IQ-2003)

Why Measuring Data Quality is Important

- Data links policy to operations i.e. it translates objectives into measurable performance indicators
- Output measurements reconcile silos of data from disparate agencies and link it to objectives
- Decisions made on fragmented or inaccurate data are invariably bad making partnerships ineffective.
- Analysis tools are useless if they do not provide accurate, measurable intelligence for action, monitoring and evaluation

4

8th International Conference on Information Quality (IQ-2003)

What happens when you can't measure data quality?

- Meta Group recently estimated there was a 50-70% failure rate across the board for SCM, ERP and CRM¹
- The Data Warehousing Institute in Seattle states data quality problems cost U.S. businesses more than \$600 billion per year²
- Ted Friedman, Principal Analyst, Gartner states he does not know a single enterprise that does not spend the last week of its business quarter cranking numbers in spreadsheets³
- The numbers might be debatable but the message is clear
 - ◆ Companies are spending more time on finding and distributing data than ensuring the data is of good enough quality to use
 - ◆ Experts have little idea of what happens to data inside IT and little idea of its fitness for purpose when it comes out

1 <http://www.meta.com/industry/industry.html?navid=19-2828>
2 Eckerson, Wayne.W. "Data Quality and the Bottom Line" TDWI Series, January 2002 <http://www.dwi-institute.com/>
3 Friedman, Ted. "Current Trends in Data Quality" June 2, 2003 http://www.infusoft.com/telecontent/webinar/SAP_entry.asp

5

8th International Conference on Information Quality (IQ-2003)

The Problem: Measuring Data Quality Outputs

Legacy Systems → Extract Transform Load → Enterprise Systems (Data Warehouse, Data Mart, Meta Data, CRM, ERP, Other) → Outputs Analysis → Knowledge Worker

Accuracy? Without evidence verification is just opinion

6

8th International Conference on Information Quality (IQ-2003)

When things goes wrong. . .

id	name	lat	lon	country	city	street	house	zip
1	530207	51.51423	0.125	GB	London	100	100	W1D 3HU
2	530208	51.51423	0.125	GB	London	100	100	W1D 3HU
3	530209	51.51423	0.125	GB	London	100	100	W1D 3HU
4	530210	51.51423	0.125	GB	London	100	100	W1D 3HU

8th International Conference on Information Quality (IQ-2002)

Barclays Bank Case Study

What did Barclays Achieve?

8th International Conference on Information Quality (IQ-2003)

What did Barclays Achieve?

- Problem
- Goals
- Requirements
- Results
- Other Applications which emerged

8th International Conference on Information Quality (IQ-2003)

Barclays Goals & Requirement

The Problem

How to generate an accurate map of customers across a large number of systems without necessarily implementing costly change at the operational level to more easily test designs ahead of emerging Basel Accord II requirements.

Goals	Requirements
<ul style="list-style-type: none"> ■ Rationalise existing data maps ■ Facilitate Basel II compliance through improved audit trails ■ Maximise existing IT investment ■ Maximise capital efficiency and RORA/ROE 	<ul style="list-style-type: none"> ■ Link millions of records from multiple systems in different countries ■ Sub field audit trails to prove record accuracy ■ Existing IT <ul style="list-style-type: none"> • Work with any system anywhere without disruption • No scrap/rework/new investment • Solution: turnkey, automated, no expertise to run • Intelligence created must prove ROI on existing IT ■ Cross selling/customer profitability/Basel II

8th International Conference on Information Quality (IQ-2003)

The Results	Other Applications
<ul style="list-style-type: none"> ■ timely accurate, single view of customers available when required ■ solution is automated and requires little expertise to run ■ process sits alongside existing IT ■ detailed audit reports generated automatically to facilitate customer analysis at single record or aggregated record level ■ business rules can be changed to reflect emerging Basel requirements ■ ability to identify and eliminate duplicate records across all data sets from the Management Information ■ exception reports are generated where further investigation might be required 	<ul style="list-style-type: none"> ■ Fixing warehouse or CRM ■ Proving ROI on existing investment ■ Basel II, Sarbanes-Oxley, IAS2005, Higgs ■ Risk reduction / what if scenario planning ■ Planning / implementing new IT ■ Balanced scorecards at individual level ■ Increased Capital efficiency ■ Cross Selling ■ Customer profitability ■ KYC and money laundering

8th International Conference on Information Quality (IQ-2002)

Barclays Bank Case Study

How did Barclays Achieve it?

8th International Conference on Information Quality (IQ-2003)

How did Barclays Achieve it?

- PROBLEM**
 - IT infrastructure is a black box
 - Few understand how it works
 - No measure of output validity
 - Is data held on the IT system good or bad?
- SOLUTION**
 - Audit trail the complete history of each record
 - Know what is happening at each part of process. IT is no longer a black box. Output is validated.

13

8th International Conference on Information Quality (IQ-2003)

Barriers to Effective Audit Trail Creation

- Data quality: without cleansing 80% of 80% = 64% reliability
- Rates of change: Each month in UK, 1m people move, 100,000 postcodes change, 60,000 businesses alter
- Complexity: each extra data match variable triggers exponential increase in complexity
- Inflexible business rules
- Insufficient audit trail detail
- Infrequent updates

Barclays Approach

- Extract data from source systems
- Create data map of entity and business relationships at sub-field level
- Build in client business rule matching criteria
- Frequently integrate source system changes with model to maintain currency

Audit Trail Potential For Use in Future Enterprise Data Projects

- Links millions of records across multiple systems without disruption
- Turnkey, automated; fits any IT, easy to use in house
- Any number of data sources added as required
- Regularly updated and easily changed to meet new reporting needs
- Test existing IT outputs and design changes

14

8th International Conference on Information Quality (IQ-2003)

The Complexity of Creating an Audit Trail

15

8th International Conference on Information Quality (IQ-2003)

FROM THE START

Legacy data... each action generates an audit trail point

FROM CAR	PETER HILTON COURT	3 6A	140 ABINGDON ST	DERBY	DE1 1DZ
AGE	PETER HILTON COURT		4 AGARD ST	DERBY	DE1 1DZ
ULT	PETER HILTON CRT		5 AGARD ST	DERBY	
LARY NON DWELLING	SIR PETER HILTON CRT				
AGE	SIR PETER HILTON CT	12A1			
F FROM CAR	SIR PETER HILTON CT	5D	AGARD ST	DERBY	
F OF CAR	SIR PETERS HILTON CT	15B	AGARD ST	DERBY	

Annotations: Missing flat numbers, No map references, Correct postcode?, 5 formats for one building, Missing street numbers, Missing postcodes.

16

8th International Conference on Information Quality (IQ-2003)

TO THE FINISH

Enterprise data...the audit trail records how the data was fixed, why decisions were made, who made them and when

3	140 ABINGDON STREET	DERBY	DE1 1DZ	PS - Z	436000	333708	
	5 AGARD STREET	DERBY	DE1 1DZ	PS - Z1Y1X	434762	336498	
	4 AGARD STREET	DERBY	DE1 1DZ	S - Z1X	434749	336495	
	5 AGARD STREET	DERBY	DE1 1RG	BS - 1Z	434648	336544	
12A1	SIR PETER HILTON COURT	AGARD STREET	DERBY	DE1 1RG	BS - C2	434648	336544
5D	SIR PETER HILTON COURT	AGARD STREET	DERBY	DE1 1RG	BS - C2	434648	336544
15B	SIR PETER HILTON COURT	AGARD STREET	DERBY	DE1 1RG	BS - C2	434648	336544

Annotations: Matched by postcode and street name, Audit trail, Matched by building and street name, One format for one building, No missing postcodes, Map references.

17

8th International Conference on Information Quality (IQ-2003)

Virtual Data Model¹ of Whole Relationship a Client has with a Bank

- A flat text file of record entity and business relationships across multiple source systems.
- Links proven by sub field, field & record audit trails. Can be used by any system to achieve a single view of the client
- All working components scripted in simple ASCII. Any application, whether Oracle, SAP or DB2, can store the model and access it with standard tools and protocols
- Performance of complex queries benefits from executing in memory and does not suffer from the read/write time delays that can be so expensive in data warehouses

¹ Hanshan, Mat. "The Virtual Data Model" August 2003 www.databis.com/gindex

18

8th International Conference on Information Quality (IQ-2003)

The Role of Artificial Intelligence in Creating a Virtual Data Model

- automating cleaning generates 50-80% accuracy; the floating 20-50% inaccuracy is the cause of data quality troubles affecting industry at the moment
- To get 90-95% accuracy, you need to work with people i.e. subject specialists, case workers, data owners et al and ask if you have the data in this system and add it to that one, what can it tell you and why does it tell you it in that way?
- Once you have done this, you are then in a position where you can build an automated script within your virtual model that mimics the way a specialist thinks and collaborates with peers in other departments
- if you focus on just data you get 'the data tells you this'. What you really need is a method of finding what you want from the data and that is what the virtual data model delivers

19

8th International Conference on Information Quality (IQ-2003)

Business Rules Matrix for Exploring the Potential of a Virtual Data Model to Answer End User Questions

Match Code	Name matching - minimum acceptable %								
	Client	Account	Trade	Credit	Loans	Foreign	Insurance	Claims	Rating
A	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
AN	60	60	60	75	75	80	70	60	85
PN	65	65	65	80	80	80	75	65	85
SN	65	70	65	80	80	85	80	65	90
AND	50	n/a	n/a	n/a	n/a	70	65	55	n/a
PND	55	n/a	n/a	n/a	n/a	70	65	65	n/a
SND	55	n/a	n/a	n/a	n/a	75	65	75	n/a
ND	75	n/a	n/a	n/a	n/a	65	65	75	n/a
ANL	n/a	n/a	n/a	n/a	65	n/a	65	65	n/a
PNL	n/a	n/a	n/a	n/a	65	n/a	65	65	n/a
SNL	70	n/a	n/a	n/a	70	n/a	70	75	n/a
NL	n/a	n/a	n/a	n/a	75	n/a	75	75	n/a
NL	n/a	n/a	n/a	n/a	70	n/a	70	70	n/a
ANL	60	n/a	50	50	50	n/a	n/a	50	50
PNI	40	n/a	50	50	50	50	n/a	50	50
SNL	40	n/a	50	50	50	50	n/a	50	50
NI	40	n/a	50	50	50	50	n/a	50	50
ACI	0	n/a	0	0	0	0	n/a	n/a	n/a
ANC	40	45	45	55	55	60	n/a	n/a	n/a
PNC	45	45	45	60	60	60	n/a	n/a	n/a
SNL	50	50	50	65	65	70	n/a	n/a	n/a
NO	80	80	80	85	85	85	n/a	n/a	n/a
NT	80	n/a	80	85	90	n/a	n/a	n/a	n/a

Virtual Data Model Matching Criteria

20

8th International Conference on Information Quality (IQ-2003)

Conclusions

- Which comes First - Data Quality or IT Design?
 - Data Quality**
Then use the outputs to plan/fine tune IT
- Why are sub field audit trails important?
 - Hard Evidence**
You can only prove output quality or establish a baseline for measuring ROI with hard evidence such as that provided by a sub field audit trail
- What is the role of Artificial Intelligence?
 - Complexity**
Traditional IT cannot cope ever changing complex variables linking records or the infinite range of end user queries

21